



CT/AU98/01023

091581286

REC'D 12 JAN 1999

WIPO PCT

Patent Office
Canberra

5

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES, hereby certify that the annexed is a true copy of the Provisional specification in connection with Application No. PP 3338 for a patent by CSL LIMITED filed on 5 May 1998.

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



WITNESS my hand this Fifth
day of January 1999

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND
SALES

AUSTRALIA

Patents Act 1990

CSL LIMITED

PROVISIONAL SPECIFICATION

Invention Title:

P. gingivalis sequences

The invention is described in the following statement:

P. gingivalis sequences

FIELD OF THE INVENTION

5 The present invention relates to *P. gingivalis* nucleotide sequences, *P. gingivalis* polypeptides and probes for detection of *P. gingivalis*.

BACKGROUND OF THE INVENTION

10 Periodontal diseases are bacterial-associated inflammatory diseases of the supporting tissues of the teeth and range from the relatively mild form of gingivitis, the non-specific, reversible inflammation of gingival tissue to the more aggressive forms of periodontitis which are characterised by the destruction of the tooth's supporting structures. Periodontitis is associated
15 with a subgingival infection of a consortium of specific Gram-negative bacteria that leads to the destruction of the periodontium and is a major public health problem. One bacterium that has attracted considerable interest is *P. gingivalis* as the recovery of this microorganism from adult periodontitis lesions can be up to 50% of the subgingival anaerobically
20 cultivable flora, whereas *P. gingivalis* is rarely recovered, and then in low numbers, from healthy sites. A proportional increase in the level of *P. gingivalis* in subgingival plaque has been associated with an increased severity of periodontitis and eradication of the microorganism from the cultivable subgingival microbial population is accompanied by resolution of
25 the disease. The progression of periodontitis lesions in non-human primates has been demonstrated with the subgingival implantation of *P. gingivalis*. These findings in both animals and humans suggest a major role for *P. gingivalis* in the development of adult periodontitis.

30 *P. gingivalis* is a black-pigmented, anaerobic, asaccharolytic, proteolytic Gram-negative rod that obtains energy from the metabolism of

specific amino acids. The microorganism has an absolute growth requirement for iron, preferentially in the form of haeme or its Fe(III) oxidation product haemin and when grown under conditions of excess haemin is highly virulent in experimental animals. A number of virulence factors have been implicated in the pathogenicity of *P. gingivalis* including the capsule, adhesins, cytotoxins and extracellular hydrolytic enzymes.

In order to develop an efficacious and safe vaccine to prevent *P. gingivalis* colonisation it is necessary to identify and produce antigens that are involved in virulence that have utility as immunogens to generate neutralising antibodies. Whilst it is possible to attempt to isolate antigens directly from cultures of *P. gingivalis* this is often difficult. For example as mentioned above, *P. gingivalis* is a strict anaerobe and can be difficult to isolate and grow. It is also known that, for a number of organisms, when cultured *in vitro* that many virulence genes are down regulated and the encoded proteins are no longer expressed. If conventional chemistry techniques were applied to purify vaccine candidates potentially important (protective) molecules may not be identified. With DNA sequencing, as the gene is present (but not transcribed) even when the organism is grown *in vitro* it can be identified, cloned and produced as a recombinant DNA protein. Similarly, a protective antigen or therapeutic target may be transiently expressed by the organism *in vitro* or produced in low levels making the identification of these molecules extremely difficult by conventional methods.

With serological identification of therapeutic targets one is limited to those responses which are detectable using standard methods such as Western Blotting or ELISA. The limitation here is the both the level of response that is generated by the animal or human and determining whether this response is protective, damaging or irrelevant. No such limitation is present with a sequencing approach to the identification of potential therapeutic or prophylactic targets.

It is also well known that *P. gingivalis* produces a range of broadly active proteases (University of Melbourne International Patent Application No PCT /AU 96/00673, US Patent Nos 5,475,097 and 5,523,390), which make the identification of intact proteins difficult because of their degradation by 5 these proteases.

SUMMARY OF THE INVENTION

The present inventors have attempted to isolate *P. gingivalis* 10 nucleotide sequences which can be used for recombinant production of *P. gingivalis* polypeptides and to develop nucleotide probes specific for *P. gingivalis*. The DNA sequences listed below have been selected from a large number of *P. gingivalis* sequences according to their indicative potential as vaccine candidates. This intuitive step involved comparison of 15 the deduced protein sequence from the *P. gingivalis* DNA sequences to the known protein sequence databases. Some of the characteristics used to select useful vaccine candidates include; the expected cellular location, such as outer membrane proteins or secreted proteins, particular functional activities of similar proteins such as those with an enzymatic or proteolytic 20 activity, proteins involved in essential metabolic pathways that when inactivated or blocked may be deleterious or lethal to the organism, proteins that might be expected to play a role in the pathogenesis of the organism eg. red cell lysis, cell agglutination or cell receptors and proteins which are paralogues to proteins with proven vaccine efficacy. DNA sequences that 25 were considered to be poor vaccine candidates and not selected include those that code for proteins involved in replication, non-essential proteins involved in cellular processes and those proteins present at sites that would be unlikely to be affected by immune mediators such as those found in the bacterial cytoplasm or inner membranes.

In a first aspect the present invention consists in an isolated *P. gingivalis* nucleotide sequence, the nucleotide sequence consisting of or including a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, 5 fragments thereof and sequences complementary thereto.

In a second aspect the present invention consists in an isolated *P. gingivalis* polypeptide, the polypeptide being at least partially encoded by a nucleotide consisting of or including a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, 10 SEQ ID NO: 5, SEQ ID NO: 6, fragments thereof and sequences complementary thereto.

In a third aspect the present invention consists in an isolated *P. gingivalis* polypeptide consisting of or including an amino acid sequence selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID 15 NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18 or fragments thereof.

In a fourth aspect the present invention consists in an isolated polynucleotide, the polynucleotide encoding a polypeptide consisting of or 20 including an amino acid sequence selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18 or fragments thereof.

In a fourth aspect the present invention consists in a nucleotide probe 25 specific for *P. gingivalis*, the probe including a detectable label and a nucleotide sequence of at least 15 nucleotides, the nucleotide sequence being derived from a sequence selected from the group consisting SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6 and sequences complementary thereto.

In a fifth aspect the present invention consists in a composition for use in raising an immune response in an animal directed against *P.*

Gingivalis the composition including an acceptable carrier and/or adjuvant and at least one polypeptide having a sequence selected from the group 5 consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18 or fragments thereof.

As will be understood by those skilled in the art the nucleotides of the 10 present invention may be useful in DNA vaccination to reduce the incidence and/or severity of *P. gingivalis* infection.

Accordingly in a sixth aspect the present invention consists in a composition for use in inducing an immune response, the composition including at least one DNA molecule, the at least one DNA molecule having or including a sequence selected from the group consisting of SEQ ID NO: 1, 15 SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, fragments thereof and sequences complementary thereto.

Further information regarding DNA vaccination may be found in Donnelly *et al*, Journal of Immunological Methods 176(1994) 145-152, the disclosure of which is incorporated herein by reference.

20 Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

25

DETAILED DESCRIPTION

Preparation of the *P. gingivalis* library for sequencing.

To determine the DNA sequence of *P. gingivalis* genomic DNA was isolated from *P. gingivalis* strain W50 (ATCC 53978) essentially by the method described by Mamur J. (1961)(1). Cloning of DNA fragments was performed essentially as described by Fleischmann *et al.*, (1995)(2). Briefly, purified genomic DNA from *P. gingivalis* was nebulized to fragment the DNA and was treated with Bal31 nuclease to create blunt ends then run twice on preparative 1% agarose gels. DNA fragments of 1.6-2.0 kb were excised from the gel and the DNA recovered. This DNA was then ligated to the vector pUC18 (*Sma*I digested and dephosphorylated; Pharmacia) and electrophoresed on a 1% agarose preparative gel. The fragment comprising linear vector plus one insert was excised, purified and this process repeated to reduce any vector without insert contamination. The recovered vector plus insert DNA was blunt-ended with T4 DNA polymerase, then a final ligation to produce circular DNA was performed. Aliquots of Epicurian Coli Electroporation-Competent Cells (Stratagene) were transformed with the library DNA and plated out on SOB agar antibiotic diffusion plates containing X-gal and incubated at 37°C overnight. Colonies with inserts appeared white and those without inserts (vector alone) appeared blue. Plates were stored at 4°C until the white clones were picked and expanded for the extraction of plasmid DNA for sequencing.

DNA sequencing

Plasmid DNA was prepared by picking bacterial colonies into 1.5ml of LB, TB or SOB broth supplemented with 50-100ug/ml Ampicillin in 96 deep well plates. Plasmid DNA was isolated using the QIAprep Spin or QIAprep 96 Turbo miniprep kits (QIAGEN GmbH, Germany). DNA was eluted into a 96 well gridded array and stored at -20C.

Sequencing reactions were performed using ABI PRISM Dye Terminator and ABI PRISM BIGDye Terminator Cycle Sequencing Ready Reaction kits with AmpliTaq DNA polymerase FS (PE Applied Biosystems,

Foster City, CA) using the M13 Universal forward and reverse sequencing primers. Sequence reactions were conducted on either a Perkin-Elmer GeneAmp 9700 (PE Applied Biosystems) or Hybaid PCR Express (Hybaid, UK) thermal cyclers. Sequencing reactions were analysed on ABI PRISM 5 377 DNA sequencers (PE Applied Biosystems).

The sequences obtained are set out below.

SeqID#1-9 represents the DNA sequence that encodes proteins in SeqID#10-27. Proteins in SeqID#10-18 are the entire open reading frame from DNA SeqID#1-9. Proteins in SeqID#19-27 are the proteins encoded by 10 DNA SeqID#1-9 from their putative initiation codon. The initiation codon was calculated from sequence homology alignment using FastX or by the 15 ORF prediction program GeneMark.

As will be understood by those skilled in the art open reading frames (ORFs) may be readily identified. ORFs may be determined using two 15 methods, for example, alignments from FastX search results may be used to define the start and end positions of coding regions if sufficient protein homology is present. Alternatively, protein coding regions may be identified using the ORF identification program GeneMark (3) using a training matrix based on published *P.gingivalis* sequences. This matrix may be further 20 refined by adding ORFs identified from the results of homology searching and ORFs identified by GeneMark. The program PSORT (4) may be used for the detection of signal sequences and the prediction of protein cell localisation. A UNIX version of TopPred (5) may be used to identify 25 potential membrane spanning domains.

25

DNA sequence analysis

Raw trace data files from the ABI 377 sequencer were manually 30 trimmed using Staden Pregap (Laboratory of Molecular Biology, Medical Research Council, UK) running on a Sun Microsystems computer. Trimmed files were assembled into contigs using Staden Gap v4.1 and exported as a

consensus file in FastA format. This consensus was converted into GCG format files and analysed for homology using the FASTX algorithm on a non-redundant protein database compiled by ANGIS (Australian Genomic Information Service, University of Sydney). Individual FASTX search results were examined for significant homology by statistical probability and amino acid alignments.

5 The results are set out in Table 1.

SeqID#	Length of protein in SeqID	Homology description	Length of protein homolog	% identity	Overlap	E value
1, 7, 13	307aa	Amylovoran outer membrane export protein, <i>Erwinia amylovora</i>	377aa	29	251aa	7.00E-06
2, 8, 14	429aa	48kD outer membrane protein, <i>Actinobacillus pleuropneumoniae</i>	449aa	32	425aa	5.70E-43
3, 9, 15	315aa	Adhesin protein, <i>Synechocystis</i> sp.	338aa	31	294aa	8.10E-14
4, 10, 16	331aa	36kD outer membrane protein, <i>Helicobacter pylori</i>	329aa	37	326aa	5.50E-43
5, 11, 17	223aa	Outer membrane porin F, <i>Pseudomonas fluorescens</i>	326aa	34	199aa	1.10E-11
6, 12, 18	426aa	Haeme uptake protein A, <i>Bacteroides fragilis</i>	431aa	80	411	9.50E-153

Table 1

(2) INFORMATION FOR SEQ ID NO:1

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 960 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: *misc_feature*
- (B) LOCATION 1...960

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1

ccgcaatcct	ctcctgatcg	aagaagcttc	caaaacgtca	tgaataaaata	tcattctcaa	60
agcgttttag	aggtcggcaa	aattgggatt	gtgattatct	ttgcgcctat	agtacggaat	120
gtacatcaac	aaccccttt	ttaagccat	aaatcaatta	tgcgtattgt	cagtaatttt	180
ttgttcgtct	cttttcgtt	tttgctttt	gcatcatgcc	gttcccagcg	agaaaaaggtc	240
gtttacactgc	aagatatacca	aacttttaat	cgggagattt	tcgctaaacc	atatgacgta	300
aaaattgaga	aggacgatgt	gtcgaacatc	cttgcagca	gttagagaccc	ggagcttca	360
acgccttaca	accaagtgtt	gaccactcg	gcactggccc	gcaacggcta	tggAACGAAC	420
tcgaacgaag	gttccctgg	cgattcgaaa	gggtacatca	attatcctat	tttaggcccag	480
atctatgtat	aggcccttac	tcttaccgaa	ctggagaagg	agatacagaa	gaggattatt	540
tccagtggat	ttatcaagga	tcctacggta	acgggtgcgc	ttcaaaaattt	caagggtgtcg	600
gttttgggag	aggtgaatca	tccgggttcg	atgtcggtaa	aaggagagcg	aataactctt	660
ttggaagcga	tcggaatggc	cgagacctg	acaatctatg	gtcgccgcga	tcgggttttt	720
gtgattagag	aaaccgatgg	gcatcgcgag	gtttccaga	cggtatctcg	aaaggccgac	780
ttgctcgcaa	gccccgtgt	ctatctgc	cagaacgcac	tcatctatgt	ggagccgaac	840
gacaagaaaa	cacagatgag	cgagatcaac	cagaataata	acgtaaaacgt	atggctgagt	900
gttacctcca	ctttggtata	catttccacg	ctgacgatta	cgataataga	taagaccaaa	
960						

(2) INFORMATION FOR SEQ ID NO:2

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1338 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: *misc_feature*

(B) LOCATION 1...1338

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2

(2) INFORMATION FOR SEQ ID NO:3

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 987 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

ORIGINAL SOURCE: *Journal of the American Revolution*

FEATURE:
(A) NAME/KEY: misc_feature

ITEM DESCRIPTION - SEC. ID. NO.: 3

aaccttagga cacagcctt cttttggta gatttgcaat ctatgatcg aacgatactt	60
tcacgatatg ttcctcgaa cttttggagt cggggagcta cctttttttt cacgatttc	120
ccggccttca tctcgccgc tactgtttg ccggcttgtg gaggggggtac tgcttcaggc	180
tccgatcgta cgctggctgt gaccatcgag ccacagaaat acttcatcg gtccattgcg	240
gataagtccgg tgcaggtggg ggcattggta ccggccggca gcaatccgg aataacgac	300
ccttcgccta ccgtatgaa gctttgtcc gaagcagatg cctacttcta tataggagga	360
ctggggttcg agcaaagaaa ttcgtgtcc attcgggaca ataaccctaa gtccttcctt	420
ttcgaaatgg gcaaagcctt ggcggatgcc ggaagtgcag atctccacgg ctccgtcaca	480
gatcattctc atacagacct gcatgcccat gatccgcact attggagcag tgggttaggg	540
gcaaaaggcac tcagtcgtgc tgcatcagac gcgcttggta agctttatcc gaacgagaaa	600

gacaaatggg acaaaggca cgaccgtctc aacggacgt a cgacacggt gaagagactc	660
gtcgatacca tggccaa tggcaaagca gacaaggct tcgtcatata tcacccatcg	720
ctcagcttt tcgcccaga gtcggctg cggcagatcg tcatagagga agatggaaa	780
gaggctacgg ctgcccacct tcgtcggtg atcgatcagg cacgtgccga tgggtcaga	840
atcgatattt tccaaaccga attgaaacg cgtcaggcgg aggacatcgc acggagatc	900
ggtgctcgtc cggttaaggat caatcctctg cgcagctgt gggaggagga aattttacat	960
attgctcgcg ctggctca tgaacgg	987

(2) INFORMATION FOR SEQ ID NO:4

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1095 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Porphyromonas gingivalis

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...1095

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4

tgctgcgcaa ggcttcggga cacctctccc cggaaagaagt cgtttcggaa tgataccgtt	60
tttccttatt tcgttattca tctgtcaaaa catattatca ttatgcgga caacaaacaa	120
cgtaatatcg tattccggc gtttctcctc ttgctggag tcatcgcat ggtgacgatc	180
gttggttttt tcatgtcgag acggccgag gagattatcc aaggacagat agaagtgacc	240
gaataccgag tgtccagcaa agtgcgggg cgcataagg aacttaggt atccgaggga	300
cagcagggtgc aggccggcga taccctcggt gtcatcgaa ccccccacgt agccgctaag	360
atggagcagg caaaggctgc cgaaggcagct gcacaggctc agaacgccaa ggctctcaaa	420
ggagcacgca gcaacagat acaggcagcc tatgagatgt ggcagaaagc tcaggccggc	480
gtagccatag cgaccaagac acaccagcgc gtgcagaacc tctatgacca gggagtggta	540
ccggctcaga agttggacga agccactgccc cagcgcgtg cggccatcgac taccaaaaaa	600
gcccggcgaag cccagtacaa tatggctcgac aacggtgccg aacgcgaaga caagctggca	660
gcttctgccc tcgtcgatag agcgagagga gccgtcgccg aggtggagtc gtacatcaac	720
gaaaccttacc tcatgccttcc acgggcaggc gaagtgtcg agatattccc caaagccggc	780
gaactcgttag gtacggcgc acctatcatg aatatcgccg agatggcga tatgtggcc	840
agctttgcgc ttctgtggaa ttccctcagc agcatgacca tgggagccgt tctggagact	900
gtgggtccgg ctctgaatga agaaaaagta cgcttcaaga tcacattcat caagaacatg	960
ggtagctatg ctgcctggaa agcgaccaag acaacaggc agtacgaccc gaagaccttc	1020
gaggtaaagg ccacccttgc ggataaagac aaggcacaaa agctacgccc ggttatgtcc	1080
gtgatcatac gcaag	1095

(2) INFORMATION FOR SEQ ID NO:5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 693 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION 1...693

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5

acgaataaaa aagaagagac aatgaagaaa tcaagtgtag tagcctcagt tttggccgtg	60
gctctcggtgt tcgcccgttg cggactgaac aatatggcaa aaggcggcct tatcggcgcc	120
ggagtaggag gtgcattgg tgcggagta ggtaacgtag ccggaaatac ggctgtcggt	180
gccatcgctcg gtactcgagt cggtggagca gccgggtgtc tcatcgaaa gaagatggac	240
aacagaaaa aagaactgga ggccgcagta cccgatgcta cgattcagac agtaaatgac	300
ggagaggcta ttctggttac ttgcatacg ggtatcctct ttgcgcacaa ctccagcact	360
ctgatccca actcacgcac tgcgctgacg aagtttgcgt caaacatgaa caaaaacccc	420
gacacggata ttctgtatcg aggccatacg gacaataccg gctccgacaa gatcaacgat	480
cctctgtctg agagacgtgc agccagcgtt tattttcc tgaattctca ggggtgtgagt	540
atgcgcga tggcagccga agggcgtgg agccatgaac cggttcaga caatagcaca	600
gttgcggac gttcgccaa cccgcgttg gaggttata tcttgcgaa tgccaagatg	660
atcgaacaag cacagcaagg tacgctgaag taa	693

(2) INFORMATION FOR SEQ ID NO:6

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1506 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

(A) NAME/KEY: misc_feature
 (B) LOCATION 1...1506

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6

gaagggggcca gcttcgttct tgcctcacc ggcaagggtc tcgtgcggc tcgtctttgg	60
gagcgcgtga acgaggctct cggaaaagac gaaatgtaa agggctatgt gaagtgtcg	120
accaagggtg gtatgtatcg cgatgtatcc ggtatcgagg ctttcctccc gggatcacag	180
atcgacgtgc gccccattcg cgactacgtat gcattcggtt agaagacgtat ggagttcaag	240
atgtgaaaaa tcaatcaaga atataagaat gtatgtttt cccacaaggat gtcatcgaa	300
gcagagctcg aacaacagaa gaaagaaatc atccgcacgc tcgaaaaagg gcaggtactc	360
gaaggatcg tcaagaatat tacttcttac ggagtattta tcgaccccg ggaggtggat	420
ggttttatcc atatcactga ctttcattgg ggtcggtgg ctcatccgaa agaaatcgta	480
cagctggatc agaagatcaa tgcgttac ctgcactttt atgaagatcg caagcgatc	540
gctctcgac tcaaaccatc gatgcctcat cttggatc ctctcgacag cgagcttaag	600
gttggcgata aggtgaaggg taaagttgtg gtatggcag attacgggtc tttcggtgag	660

attgcacagg	gcgttggagg	tcttatccac	gtaagcggaa	tgtcatggac	acagcacttg	720
cgttctgctc	aggacttcct	gcatgttaggc	gacgaagtgg	aagccgtgat	cctgacgctc	780
gaccgcgaag	aacgcaaaat	gtcgctcggt	ctgaagcaac	tcaagccgga	tccttggct	840
gatatcgaaa	ctcgttccc	tgttaggctct	cgtcaccatg	ctcggttgc	caacttcacc	900
aatttcggtg	tattcggtga	gatcgaagag	ggcgttagatg	gccttatcca	tattccgac	960
ctttcttgg	cgaaaaagat	caaacacccc	agcgagttt	cggaagttagg	tgctgtatc	1020
gaagttcagg	taatcgagat	cgacaaggaa	aaccgtcgtc	ttagcttggg	tcacaaacag	1080
ttggaagaga	atccttggg	tgtattcgg	acggatttca	ctgtaggatc	tatccacgaa	1140
ggaacggtaa	tcgaagtgtat	ggacaagggt	gctgtcggtt	ctctgcctt	cggtgtggaa	1200
ggttttgc	ctccgaagca	catggtaag	gaagatggct	cacaggctgt	actcgaagag	1260
aagttacctt	tcaaggttat	ttagttcaat	aaggatgcca	agcgaatcat	tgtatctcat	1320
agccgtgtat	tcgaagatga	gcagaaaaatg	gctcagcgtg	aagccaatgc	agagcgtaag	1380
gctgaagcca	aagcgctca	gaaagaagct	gctgccgaag	ctgccaatcc	tgcacaggct	1440
gtagagaaaag	ccactctcg	agacctcg	gagctggccg	ctttgaaaga	aaagctttca	1500
gaaaaac						1506

(2) INFORMATION FOR SEQ ID NO:7

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 320 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...320

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7

Pro	Gln	Ser	Ser	Pro	Asp	Arg	Arg	Ser	Phe	Gln	Asn	Val	Met	Asn	Lys	
1																
														15		
Tyr	His	Ser	Gln	Ser	Val	Leu	Glu	Val	Gly	Lys	Ile	Gly	Ile	Val	Ile	
															30	
Ile	Phe	Ala	Pro	Ile	Val	Arg	Asn	Val	His	Gln	Gln	Pro	Pro	Phe	Leu	
														45		
Ser	His	Lys	Ser	Ile	Met	Arg	Ile	Val	Ser	Asn	Phe	Leu	Phe	Val	Ser	
															60	
Phe	Ser	Val	Leu	Leu	Phe	Ala	Ser	Cys	Arg	Ser	Gln	Arg	Glu	Lys	Val	
															80	
Val	Tyr	Leu	Gln	Asp	Ile	Gln	Thr	Phe	Asn	Arg	Glu	Ile	Ile	Ala	Lys	
															95	
Pro	Tyr	Asp	Val	Lys	Ile	Glu	Lys	Asp	Asp	Val	Leu	Asn	Ile	Leu	Val	
															110	
Ser	Ser	Arg	Asp	Pro	Glu	Leu	Ser	Thr	Pro	Tyr	Asn	Gln	Val	Leu	Thr	
															125	
Thr	Arg	Ala	Leu	Ala	Arg	Asn	Gly	Tyr	Gly	Thr	Asn	Ser	Asn	Glu	Gly	
															140	
Phe	Leu	Val	Asp	Ser	Lys	Gly	Tyr	Ile	Asn	Tyr	Pro	Ile	Leu	Gly	Gln	
															160	
Ile	Tyr	Val	Glu	Leu	Thr	Arg	Thr	Glu	Leu	Glu	Lys	Glu	Ile	Gln		
															175	
Lys	Arg	Ile	Ile	S	r	Ser	Gly	Phe	Ile	Lys	Asp	Pro	Thr	Val	Thr	Val
															190	
Gln	Leu	Gln	Asn	Phe	Lys	Val	Ser	Val	Leu	Gly	Glu	Val	Asn	His	Pro	

195	200	205	
Gly Ser Met Ser Val Lys	Gly Glu Arg Ile Thr Leu	Leu Glu Ala Ile	
210	215	220	
Gly Met Ala Gly Asp Leu	Thr Ile Tyr Gly Arg Arg	Asp Arg Val Phe	
225	230	235	240
Val Ile Arg Glu Thr Asp	Gly His Arg Glu Val Ph	Gln Thr Asp Leu	
245	250	255	
Arg Lys Ala Asp Leu	Leu Ala Ser Pro Val	Tyr Tyr Leu His Gln Asn	
260	265	270	
Asp Val Ile Tyr Val Glu	Pro Asn Asp Lys Lys	Thr Gln Met Ser Glu	
275	280	285	
Ile Asn Gln Asn Asn Asn	Val Asn Val Trp	Leu Ser Val Thr Ser Thr	
290	295	300	
Leu Val Ser Ile Ser Thr	Leu Thr Ile Thr Ile	Ile Asp Lys Thr Lys	
305	310	315	320

(2) INFORMATION FOR SEQ ID NO:8

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 446 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...446

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8

Asn Pro Lys Lys Ala Leu Ala Leu Asn Leu Lys Gly Lys Pro Leu Pro			
1	5	10	15
Glu Met Leu Ala Glu Pro Ala Gln Ser Pro Thr Tyr Ala Val Val Pro			
20	25	30	
Asp Asp Phe Glu Gly Val Ile Pro Lys Val Thr Ala Arg Pro Gly Asp			
35	40	45	
Lys Val Arg Ala Gly Ser Ala Leu Met His His Lys Ala Tyr Pro Glu			
50	55	60	
Met Lys Phe Thr Ser Pro Val Ser Gly Glu Val Ile Ala Val Asn Arg			
65	70	75	80
Gly Ala Lys Arg Lys Val Leu Ser Ile Glu Val Lys Pro Asp Gly Leu			
85	90	95	
Asn Glu Tyr Glu Ser Phe Pro Val Gly Asp Pro Ser Ala Leu Ser Ala			
100	105	110	
Glu Gln Ile Lys Glu Leu Leu Leu Ser Ser Gly Met Trp Gly Phe Ile			
115	120	125	
Lys Gln Arg Pro Tyr Asp Ile Val Ala Thr Pro Asp Ile Ala Pro Arg			
130	135	140	
Asp Ile Tyr Ile Thr Ala Asn Phe Thr Ala Pro Leu Ala Pro Asp Phe			
145	150	155	160
Asp Phe Ile Val Arg Gly Glu Glu Arg Ala Leu Gln Thr Ala Ile Asp			
165	170	175	
Ala Leu Ala Lys Leu Thr Thr Gly Lys Val Tyr Val Gly Leu Lys Pro			
180	185	190	
Gly Ser Ser Leu Gly Leu His Asn Ala Glu Ile Val Glu Val His Gly			

195	200	205	
Pro His Pro Ala Gly Asn Val	Gly Val Leu Ile Asn His Thr Lys Pro		
210	215	220	
Ile Asn Arg Gly Glu Thr Val Trp	Thr Leu Lys Ala Thr Asp Leu Il		
225	230	235	240
Val Ile Gly Arg Phe Leu Leu Thr	Gly Lys Ala Asp Phe Thr Arg Met		
245	250	255	
Ile Ala Met Thr Gly Ser Asp Ala	Ala His Gly Tyr Val Arg Ile		
260	265	270	
Met Pro Gly Cys Asn Val Phe Ala	Ser Phe Pro Gly Arg Leu Thr Ile		
275	280	285	
Lys Glu Ser His Glu Arg Val Ile	Asp Gly Asn Val Leu Thr Gly Lys		
290	295	300	
Lys Leu Cys Glu Lys Glu Pro Phe	Leu Ser Ala Arg Cys Asp Gln Ile		
305	310	315	320
Thr Val Ile Pro Glu Gly Asp Asp	Val Asp Glu Leu Phe Gly Trp Ala		
325	330	335	
Ala Pro Arg Leu Asp Gln Tyr Ser	Met Ser Arg Ala Tyr Phe Ser Trp		
340	345	350	
Leu Gln Gly Lys Asn Lys Glu	Tyr Val Leu Asp Ala Arg Ile Lys Gly		
355	360	365	
Gly Glu Arg Ala Met Ile Met Ser	Asn Glu Tyr Asp Arg Val Phe Pro		
370	375	380	
Met Asp Ile Tyr Pro Glu Tyr Leu	Leu Lys Ala Ile Ile Ala Phe Asp		
385	390	395	400
Ile Asp Lys Met Glu Asp Leu Gly	Ile Tyr Glu Val Ala Pro Glu Asp		
405	410	415	
Phe Ala Thr Cys Glu Phe Val Asp	Thr Ser Lys Ile Glu Leu Gln Arg		
420	425	430	
Ile Val Arg Glu Gly Leu Asp Met	Leu Tyr Lys Glu Met Asn		
435	440	445	

(2) INFORMATION FOR SEQ ID NO:9

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 329 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...329

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9

Asn Leu Arg Thr Gln Pro Phe Phe	Leu Val Asp Leu Gln Ser Met Ile		
1	5	10	15
Arg Thr Il Leu Ser Arg Tyr Val	Ser Ser Asn Phe Trp Ser Arg Gly		
20	25	30	
Ala Thr Phe Phe Ph Thr Ile Ph	Pro Ala Phe Ile Leu Ala Ala Thr		
35	40	45	
Ala Leu Pro Ala Cys Gly Gly	Thr Ala Ser Gly Ser Asp Arg Thr		
50	55	60	
Leu Ala Val Thr Ile Glu Pro Gln	Lys Tyr Phe Ile Glu Ser Ile Ala		

35 70 75 80
 Asp Lys Ser Val Gln Val Val Ala Leu Val Pro Ala Gly Ser Asn Pro
 85 90 95
 Glu Glu Tyr Asp Pro Ser Pro Thr Val Met Lys Arg Leu Ser Glu Ala
 100 105 110
 Asp Ala Tyr Phe Tyr Ile Gly Gly Leu Gly Phe Glu Gln Arg Asn Leu
 115 120 125
 Ala Ala Ile Arg Asp Asn Asn Pro Lys Leu Pro Leu Phe Glu Met Gly
 130 135 140
 Lys Ala Leu Ala Asp Ala Gly Ser Ala Asp Leu His Gly Ser Cys Thr
 145 150 155 160
 Asp His Ser His Thr Asp Leu His Ala His Asp Pro His Tyr Trp Ser
 165 170 175
 Ser Val Val Gly Ala Lys Ala Leu Ser Arg Ala Ala Tyr Asp Ala Leu
 180 185 190
 Val Glu Leu Tyr Pro Asn Glu Lys Asp Lys Trp Asp Lys Gly His Asp
 195 200 205
 Arg Leu Asn Gly Arg Ile Asp Ser Val Lys Arg Leu Val Asp Thr Met
 210 215 220
 Phe Ala Asn Gly Lys Ala Asp Lys Ala Phe Val Ile Tyr His Pro Ser
 225 230 235 240
 Leu Ser Phe Phe Ala Gln Glu Phe Gly Leu Arg Gln Ile Val Ile Glu
 245 250 255
 Glu Asp Gly Lys Glu Pro Thr Ala Ala His Leu Arg Arg Val Ile Asp
 260 265 270
 Gln Ala Arg Ala Asp Gly Val Arg Ile Val Phe Ile Gln Pro Glu Phe
 275 280 285
 Glu Thr Arg Gln Ala Glu Asp Ile Ala Arg Glu Ile Gly Ala Arg Pro
 290 295 300
 Val Arg Ile Asn Pro Leu Arg Ser Ser Trp Glu Glu Glu Ile Leu His
 305 310 315 320
 Ile Ala Arg Ala Leu Ala His Glu Arg
 325

(2) INFORMATION FOR SEQ ID NO:10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 365 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

ORIGINAL SOURCE:

(22) 2000

FEATURE:
(A) NAME/KEY: misc_feature

APPENDIX 270.22 NO. 10

```

Cys Cys Ala Arg Leu Arg Asp Thr Ser Pro Arg Lys Lys Ser Ph Arg
1 5 10 15
Asn Asp Thr Val Phe Pro Tyr Phe Val Ile His L u Ile Lys His Ile
20 25 30
Ile Ile M t Thr Asp Asn Lys Gln Arg Asn Ile Val Phe Pro Ala Phe
35 40 45
Leu Leu L u L u Gly Val Ile Ala Val Val Thr Ile Val Gly Phe Phe

```

50	55	60	
Met	Leu Arg Pro Ala Glu Glu	Ile Ile Gln Gly Gln Ile Glu Val Thr	
65	70	75	80
Glu	Tyr Arg Val Ser Ser Lys Val Pro Gly Arg	Ile Lys Glu Leu Arg	
	85	90	95
Val	Ser Glu Gly Gln Gln Val Gln Ala Gly	Asp Thr Leu Ala Val Ile	
	100	105	110
Glu	Ala Pro Asp Val Ala Ala Lys	Met Glu Gln Ala Lys Ala Ala Glu	
	115	120	125
Ala	Ala Ala Gln Ala Gln Asn Ala Lys	Ala Leu Lys Gly Ala Arg Ser	
	130	135	140
Glu	Gln Ile Gln Ala Ala Tyr Glu Met Trp	Gln Lys Ala Gln Ala Gly	
145	150	155	160
Val	Ala Ile Ala Thr Lys Thr His Gln Arg	Val Gln Asn Leu Tyr Asp	
	165	170	175
Gln	Gly Val Val Pro Ala Gln Lys	Leu Asp Glu Ala Thr Ala Gln Arg	
	180	185	190
Asp	Ala Ala Ile Ala Thr Gln Lys	Ala Ala Glu Ala Gln Tyr Asn Met	
	195	200	205
Ala	Arg Asn Gly Ala Glu Arg	Glu Asp Lys Leu Ala Ala Ser Ala Leu	
	210	215	220
Val	Asp Arg Ala Arg Gly	Ala Val Ala Glu Val Glu Ser Tyr Ile Asn	
225	230	235	240
Glu	Thr Tyr Leu Ile Ala Pro Arg Ala Gly	Glu Val Ser Glu Ile Phe	
	245	250	255
Pro	Lys Ala Gly Glu Leu Val Gly	Thr Gly Ala Pro Ile Met Asn Ile	
	260	265	270
Ala	Glu Met Gly Asp Met Trp	Ala Ser Phe Ala Val Arg Glu Asp Phe	
	275	280	285
Leu	Ser Ser Met Thr Met Gly	Ala Val Leu Glu Thr Val Val Pro Ala	
	290	295	300
Leu	Asn Glu Glu Lys Val Arg Phe Lys	Ile Thr Phe Ile Lys Asn Met	
305	310	315	320
Gly	Thr Tyr Ala Ala Trp Lys Ala Thr Lys	Thr Thr Gly Gln Tyr Asp	
	325	330	335
Leu	Lys Thr Phe Glu Val Lys Ala Thr	Leu Ala Asp Lys Asp Lys Ala	
	340	345	350
Gln	Lys Leu Arg Pro Gly Met Ser Val Ile Ile Arg Lys		
	355	360	365

(2) INFORMATION FOR SEQ ID NO:11

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 230 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_f atur
- (B) LOCATION 1...230

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11

Thr Asn Lys Lys Glu Glu Thr Met Lys Lys Ser Ser Val Val Ala Ser

5	10	15	
Val Leu Ala Val Ala Leu Val Phe Ala Gly Cys Gly Leu Asn Asn Met			
20	25	30	
Ala Lys Gly Gly Leu Ile Gly Ala Gly Val Gly Gly Ala Ile Gly Ala			
35	40	45	
Gly Val Gly Asn Val Ala Gly Asn Thr Ala Val Gly Ala Ile Val Gly			
50	55	60	
Thr Ala Val Gly Gly Ala Ala Gly Ala Leu Ile Gly Lys Lys Met Asp			
65	70	75	80
Lys Gln Lys Lys Glu Leu Glu Ala Ala Val Pro Asp Ala Thr Ile Gln			
85	90	95	
Thr Val Asn Asp Gly Glu Ala Ile Leu Val Thr Phe Asp Ser Gly Ile			
100	105	110	
Leu Phe Ala Thr Asn Ser Ser Thr Leu Ser Pro Asn Ser Arg Thr Ala			
115	120	125	
Leu Thr Lys Phe Ala Ala Asn Met Asn Lys Asn Pro Asp Thr Asp Ile			
130	135	140	
Arg Ile Val Gly His Thr Asp Asn Thr Gly Ser Asp Lys Ile Asn Asp			
145	150	155	160
Pro Leu Ser Glu Arg Arg Ala Ala Ser Val Tyr Ser Phe Leu Asn Ser			
165	170	175	
Gln Gly Val Ser Met Ser Arg Met Ala Ala Glu Gly Arg Gly Ser His			
180	185	190	
Glu Pro Val Ala Asp Asn Ser Thr Val Ala Gly Arg Ser Ala Asn Arg			
195	200	205	
Arg Val Glu Val Tyr Ile Leu Pro Asn Ala Lys Met Ile Glu Gln Ala			
210	215	220	
Gln Gln Gly Thr Leu Lys			
225	230		

(2) INFORMATION FOR SEQ ID NO:12

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 502 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...502

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12

Glu Gly Ala Ser Phe Val Leu Val Leu Thr Gly Lys Gly Phe Val Pro			
1	5	10	15
Ala Arg Ser Trp Glu Arg Val Asn Glu Ala Leu Glu Lys Asp Glu Ile			
20	25	30	
Val Lys Gly Tyr Val Lys Cys Arg Thr Lys Gly Gly Met Ile Val Asp			
35	40	45	
Val Phe Gly Ile Glu Ala Phe Leu Pro Gly Ser Gln Ile Asp Val Arg			
50	55	60	
Pro Ile Arg Asp Tyr Asp Ala Phe Val Glu Lys Thr Met Glu Phe Lys			
65	70	75	80
Ile Val Lys Ile Asn Gln Glu Tyr Lys Asn Val Val Ser His Lys			

85	90	95
Val Leu Ile Glu Ala Glu Leu Glu Gln Gln Lys Lys Glu	Ile Ile Gly	
100	105	110
Lys Leu Glu Lys Gly Gln Val Leu Glu Gly Ile Val Lys	Asn Ile Thr	
115	120	125
Ser Tyr Gly Val Phe Ile Asp Leu Gly Gly Val Asp Gly	Leu Ile His	
130	135	140
Ile Thr Asp Leu Ser Trp Gly Arg Val Ala His Pro Glu	Glu Ile Val	
145	150	155
Gln Leu Asp Gln Lys Ile Asn Val Val Ile Leu Asp Phe	Asp Glu Asp	
165	170	175
Arg Lys Arg Ile Ala Leu Gly Leu Lys Gln Leu Met Pro	His Pro Trp	
180	185	190
Asp Ala Leu Asp Ser Glu Leu Lys Val Gly Asp Lys Val	Lys Gly Lys	
195	200	205
Val Val Val Met Ala Asp Tyr Gly Ala Phe Val Glu	Ile Ala Gln Gly	
210	215	220
Val Glu Gly Leu Ile His Val Ser Glu Met Ser Trp Thr	Gln His Leu	
225	230	235
Arg Ser Ala Gln Asp Phe Leu His Val Gly Asp Glu Val	Glu Ala Val	
245	250	255
Ile Leu Thr Leu Asp Arg Glu Glu Arg Lys Met Ser Leu	Gly Leu Lys	
260	265	270
Gln Leu Lys Pro Asp Pro Trp Ala Asp Ile Glu Thr Arg	Phe Pro Val	
275	280	285
Gly Ser Arg His His Ala Arg Val Arg Asn Phe Thr Asn	Phe Gly Val	
290	295	300
Phe Val Glu Ile Glu Glu Gly Val Asp Gly Leu Ile His	Ile Ser Asp	
305	310	315
Leu Ser Trp Thr Lys Lys Ile Lys His Pro Ser Glu Phe	Thr Glu Val	
325	330	335
Gly Ala Asp Ile Glu Val Gln Val Ile Glu Ile Asp Lys	Glu Asn Arg	
340	345	350
Arg Leu Ser Leu Gly His Lys Gln Leu Glu Glu Asn Pro	Trp Asp Val	
355	360	365
Phe Glu Thr Val Phe Thr Val Gly Ser Ile His Glu Gly	Thr Val Ile	
370	375	380
Glu Val Met Asp Lys Gly Ala Val Val Ser Leu Pro Tyr	Gly Val Glu	
385	390	395
Gly Phe Ala Thr Pro Lys His Met Val Lys Glu Asp Gly	Ser Gln Ala	
405	410	415
Val Leu Glu Glu Lys Leu Pro Phe Lys Val Ile Glu Phe	Asn Lys Asp	
420	425	430
Ala Lys Arg Ile Ile Val Ser His Ser Arg Val Phe Glu	Asp Glu Gln	
435	440	445
Lys Met Ala Gln Arg Glu Ala Asn Ala Glu Arg Lys	Ala Glu Ala Lys	
450	455	460
Ala Ala Gln Lys Glu Ala Ala Ala Glu Ala Ala Asn	Pro Ala Gln Ala	
465	470	475
Val Glu Lys Ala Thr Leu Gly Asp Leu Glu Glu Leu Ala	Ala Leu Lys	
485	490	495
Glu Lys Leu Ser Glu Asn		
500		

(2) INFORMATION FOR SEQ ID NO:13

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 307 amino acids
- (B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION 1...307

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13

Met	Asn	Lys	Tyr	His	Ser	Gln	Ser	Val	Leu	Glu	Val	Gly	Lys	Ile	Gly
1															15
Ile	Val	Ile	Ile	Phe	Ala	Pro	Ile	Val	Arg	Asn	Val	His	Gln	Gln	Pro
															30
Pro	Phe	Leu	Ser	His	Lys	Ser	Ile	Met	Arg	Ile	Val	Ser	Asn	Phe	Leu
															45
Phe	Val	Ser	Phe	Ser	Val	Leu	Leu	Phe	Ala	Ser	Cys	Arg	Ser	Gln	Arg
															60
50															
Glu	Lys	Val	Val	Tyr	Leu	Gln	Asp	Ile	Gln	Thr	Phe	Asn	Arg	Glu	Ile
65															80
Ile	Ala	Lys	Pro	Tyr	Asp	Val	Lys	Ile	Glu	Lys	Asp	Asp	Val	Leu	Asn
															95
Ile	Leu	Val	Ser	Ser	Arg	Asp	Pro	Glu	Leu	Ser	Thr	Pro	Tyr	Asn	Gln
															110
Val	Leu	Thr	Thr	Arg	Ala	Leu	Ala	Arg	Asn	Gly	Tyr	Gly	Thr	Asn	Ser
															125
Asn	Glu	Gly	Phe	Leu	Val	Asp	Ser	Lys	Gly	Tyr	Ile	Asn	Tyr	Pro	Ile
															140
Leu	Gly	Gln	Ile	Tyr	Val	Glu	Gly	Leu	Thr	Arg	Thr	Glu	Leu	Glu	Lys
145															160
Glu	Ile	Gln	Lys	Arg	Ile	Ile	Ser	Ser	Gly	Phe	Ile	Lys	Asp	Pro	Thr
															175
Val	Thr	Val	Gln	Leu	Gln	Asn	Phe	Lys	Val	Ser	Val	Leu	Gly	Glu	Val
															190
Asn	His	Pro	Gly	Ser	Met	Ser	Val	Lys	Gly	Glu	Arg	Ile	Thr	Leu	Leu
															205
Glu	Ala	Ile	Gly	Met	Ala	Gly	Asp	Leu	Thr	Ile	Tyr	Gly	Arg	Arg	Asp
															220
Arg	Val	Phe	Val	Ile	Arg	Glu	Thr	Asp	Gly	His	Arg	Glu	Val	Phe	Gln
225															240
Thr	Asp	Leu	Arg	Lys	Ala	Asp	Leu	Leu	Ala	Ser	Pro	Val	Tyr	Tyr	Leu
															255
His	Gln	Asn	Asp	Val	Ile	Tyr	Val	Glu	Pro	Asn	Asp	Lys	Lys	Thr	Gln
															270
Met	Ser	Glu	Ile	Asn	Gln	Asn	Asn	Val	Asn	Val	Trp	Leu	Ser	Val	
															285
Thr	Ser	Thr	Leu	Val	Ser	Ile	Ser	Thr	Leu	Thr	Ile	Thr	Ile	Ile	Asp
															290
Lys	Thr	Lys													295
															300
305															

(2) INFORMATION FOR SEQ ID NO:14

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 429 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION 1...429

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14

Met	Leu	Ala	Glu	Pro	Ala	Gln	Ser	Pro	Thr	Tyr	Ala	Val	Val	Pro	Asp
1			5						10					15	
Asp	Phe	Glu	Gly	Val	Ile	Pro	Lys	Val	Thr	Ala	Arg	Pro	Gly	Asp	Lys
									25					30	
Val	Arg	Ala	Gly	Ser	Ala	Leu	Met	His	His	Lys	Ala	Tyr	Pro	Glu	Met
								35	40			45			
Lys	Phe	Thr	Ser	Pro	Val	Ser	Gly	Glu	Val	Ile	Ala	Val	Asn	Arg	Gly
							50	55			60				
Ala	Lys	Arg	Lys	Val	Leu	Ser	Ile	Glu	Val	Lys	Pro	Asp	Gly	Leu	Asn
							65	70		75			80		
Glu	Tyr	Glu	Ser	Phe	Pro	Val	Gly	Asp	Pro	Ser	Ala	Leu	Ser	Ala	Glu
							85		90				95		
Gln	Ile	Lys	Glu	Leu	Leu	Ser	Ser	Gly	Met	Trp	Gly	Phe	Ile	Lys	
							100	105			110				
Gln	Arg	Pro	Tyr	Asp	Ile	Val	Ala	Thr	Pro	Asp	Ile	Ala	Pro	Arg	Asp
							115	120			125				
Ile	Tyr	Ile	Thr	Ala	Asn	Phe	Thr	Ala	Pro	Leu	Ala	Pro	Asp	Phe	Asp
							130	135			140				
Phe	Ile	Val	Arg	Gly	Glu	Arg	Ala	Leu	Gln	Thr	Ala	Ile	Asp	Ala	
							145	150		155			160		
Leu	Ala	Lys	Leu	Thr	Thr	Gly	Lys	Val	Tyr	Val	Gly	Leu	Lys	Pro	Gly
							165		170			175			
Ser	Ser	Leu	Gly	Leu	His	Asn	Ala	Glu	Ile	Val	Glu	Val	His	Gly	Pro
							180		185			190			
His	Pro	Ala	Gly	Asn	Val	Gly	Val	Leu	Ile	Asn	His	Thr	Lys	Pro	Ile
							195	200			205				
Asn	Arg	Gly	Glu	Thr	Val	Trp	Thr	Leu	Lys	Ala	Thr	Asp	Leu	Ile	Val
							210	215			220				
Ile	Gly	Arg	Phe	Leu	Leu	Thr	Gly	Lys	Ala	Asp	Phe	Thr	Arg	Met	Ile
							225	230		235			240		
Ala	Met	Thr	Gly	Ser	Asp	Ala	Ala	Ala	His	Gly	Tyr	Val	Arg	Ile	Met
							245		250			255			
Pro	Gly	Cys	Asn	Val	Phe	Ala	Ser	Phe	Pro	Gly	Arg	Leu	Thr	Ile	Lys
							260		265			270			
Glu	Ser	His	Glu	Arg	Val	Ile	Asp	Gly	Asn	Val	Leu	Thr	Gly	Lys	Lys
							275		280			285			
Leu	Cys	Glu	Lys	Glu	Pro	Phe	Leu	Ser	Ala	Arg	Cys	Asp	Gln	Ile	Thr
							290	295			300				
Val	Ile	Pro	Glu	Gly	Asp	Asp	Val	Asp	Glu	Leu	Phe	Gly	Trp	Ala	Ala
							305	310		315			320		
Pro	Arg	Leu	Asp	Gln	Tyr	Ser	Met	Ser	Arg	Ala	Tyr	Phe	Ser	Trp	Leu
							325		330			335			
Gln	Gly	Lys	Asn	Lys	Glu	Tyr	Val	Leu	Asp	Ala	Arg	Ile	Lys	Gly	Gly
							340		345			350			
Glu	Arg	Ala	Met	Ile	Met	Ser	Asn	Glu	Tyr	Asp	Arg	Val	Phe	Pro	Met
							355		360			365			
Asp	Ile	Tyr	Pro	Glu	Tyr	Leu	Leu	Lys	Ala	Ile	Ile	Ala	Phe	Asp	Ile

370	375	380
Asp Lys Met Glu Asp Leu	Gly Ile Tyr Glu Val	Ala Pro Glu Asp Phe
	390	400
385 Ala Thr Cys Glu Phe Val Asp Thr Ser	Lys Ile Glu Leu Gln Arg Ile	
	405	415
Val Arg Glu Gly Leu Asp Met Leu Tyr	Lys Glu Met Asn	
	420	425

(2) INFORMATION FOR SEQ ID NO:15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 315 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...315

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15

Met Ile Arg Thr Ile Leu Ser Arg Tyr Val Ser Ser Asn Phe Trp Ser			
1	5	10	15
Arg Gly Ala Thr Phe Phe Phe Thr Ile Phe Pro Ala Phe Ile Leu Ala			
20	25	30	
Ala Thr Ala Leu Pro Ala Cys Gly Gly Gly Thr Ala Ser Gly Ser Asp			
35	40	45	
Arg Thr Leu Ala Val Thr Ile Glu Pro Gln Lys Tyr Phe Ile Glu Ser			
50	55	60	
Ile Ala Asp Lys Ser Val Gln Val Val Ala Leu Val Pro Ala Gly Ser			
65	70	75	80
Asn Pro Glu Glu Tyr Asp Pro Ser Pro Thr Val Met Lys Arg Leu Ser			
85	90	95	
Glu Ala Asp Ala Tyr Phe Tyr Ile Gly Gly Leu Gly Phe Glu Gln Arg			
100	105	110	
Asn Leu Ala Ala Ile Arg Asp Asn Asn Pro Lys Leu Pro Leu Phe Glu			
115	120	125	
Met Gly Lys Ala Leu Ala Asp Ala Gly Ser Ala Asp Leu His Gly Ser			
130	135	140	
Cys Thr Asp His Ser His Thr Asp Leu His Ala His Asp Pro His Tyr			
145	150	155	160
Trp Ser Ser Val Val Gly Ala Lys Ala Leu Ser Arg Ala Ala Tyr Asp			
165	170	175	
Ala Leu Val Glu Leu Tyr Pro Asn Glu Lys Asp Lys Trp Asp Lys Gly			
180	185	190	
His Asp Arg Leu Asn Gly Arg Ile Asp Ser Val Lys Arg Leu Val Asp			
195	200	205	
Thr Met Phe Ala Asn Gly Lys Ala Asp Lys Ala Phe Val Ile Tyr His			
210	215	220	
Pro S r L u Ser Phe Phe Ala Gln Glu Ph Gly Leu Arg Gln Ile Val			
225	230	235	240
Ile Glu Glu Asp Gly Lys Glu Pro Thr Ala Ala His Leu Arg Arg Val			
245	250	255	
Il Asp Gln Ala Arg Ala Asp Gly Val Arg Ile Val Phe Ile Gln Pro			

260	265	270
Glu Phe Glu Thr Arg Gln Ala Glu Asp Ile Ala Arg Glu Ile Gly Ala		
275	280	285
Arg Pro Val Arg Ile Asn Pro Leu Arg Ser Ser Trp Glu Glu Ile		
290	295	300
Leu His Ile Ala Arg Ala Leu Ala His Glu Arg		
305	310	315

(2) INFORMATION FOR SEQ ID NO:16

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 331 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...331

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16

Met Thr Asp Asn Lys Gln Arg Asn Ile Val Phe Pro Ala Phe Leu Leu		
1	5	10 15
Leu Leu Gly Val Ile Ala Val Val Thr Ile Val Gly Phe Phe Met Leu		
20	25	30
Arg Pro Ala Glu Glu Ile Ile Gln Gly Gln Ile Glu Val Thr Glu Tyr		
35	40	45
Arg Val Ser Ser Lys Val Pro Gly Arg Ile Lys Glu Leu Arg Val Ser		
50	55	60
Glu Gly Gln Gln Val Gln Ala Gly Asp Thr Leu Ala Val Ile Glu Ala		
65	70	75 80
Pro Asp Val Ala Ala Lys Met Glu Gln Ala Lys Ala Ala Glu Ala Ala		
85	90	95
Ala Gln Ala Gln Asn Ala Lys Ala Leu Lys Gly Ala Arg Ser Glu Gln		
100	105	110
Ile Gln Ala Ala Tyr Glu Met Trp Gln Lys Ala Gln Ala Gly Val Ala		
115	120	125
Ile Ala Thr Lys Thr His Gln Arg Val Gln Asn Leu Tyr Asp Gln Gly		
130	135	140
Val Val Pro Ala Gln Lys Leu Asp Glu Ala Thr Ala Gln Arg Asp Ala		
145	150	155 160
Ala Ile Ala Thr Gln Lys Ala Ala Glu Ala Gln Tyr Asn Met Ala Arg		
165	170	175
Asn Gly Ala Glu Arg Glu Asp Lys Leu Ala Ala Ser Ala Leu Val Asp		
180	185	190
Arg Ala Arg Gly Ala Val Ala Glu Val Glu Ser Tyr Ile Asn Glu Thr		
195	200	205
Tyr Leu Ile Ala Pro Arg Ala Gly Glu Val Ser Glu Ile Phe Pro Lys		
210	215	220
Ala Gly Glu Leu Val Gly Thr Gly Ala Pro Ile Met Asn Ile Ala Glu		
225	230	235 240
Met Gly Asp Met Trp Ala Ser Phe Ala Val Arg Glu Asp Phe Leu Ser		
245	250	255
Ser Met Thr Met Gly Ala Val Leu Glu Thr Val Val Pro Ala Leu Asn		

260	265	270
Glu Glu Lys Val Arg Phe Lys Ile Thr Phe Ile Lys Asn Met Gly Thr		
275	280	285
Tyr Ala Ala Trp Lys Ala Thr Lys Thr Thr Gly Gln Tyr Asp Leu Lys		
290	295	300
Thr Phe Glu Val Lys Ala Thr Leu Ala Asp Lys Asp Lys Ala Gln Lys		
305	310	315
Leu Arg Pro Gly Met Ser Val Ile Ile Arg Lys		
325	330	

(2) INFORMATION FOR SEQ ID NO:17

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 223 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...223

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17

Met Lys Lys Ser Ser Val Val Ala Ser Val Leu Ala Val Ala Leu Val			
1	5	10	15
Phe Ala Gly Cys Gly Leu Asn Asn Met Ala Lys Gly Gly Leu Ile Gly			
20	25	30	
Ala Gly Val Gly Gly Ala Ile Gly Ala Gly Val Gly Asn Val Ala Gly			
35	40	45	
Asn Thr Ala Val Gly Ala Ile Val Gly Thr Ala Val Gly Gly Ala Ala			
50	55	60	
Gly Ala Leu Ile Gly Lys Lys Met Asp Lys Gln Lys Lys Glu Leu Glu			
65	70	75	80
Ala Ala Val Pro Asp Ala Thr Ile Gln Thr Val Asn Asp Gly Glu Ala			
85	90	95	
Ile Leu Val Thr Phe Asp Ser Gly Ile Leu Phe Ala Thr Asn Ser Ser			
100	105	110	
Thr Leu Ser Pro Asn Ser Arg Thr Ala Leu Thr Lys Phe Ala Ala Asn			
115	120	125	
Met Asn Lys Asn Pro Asp Thr Asp Ile Arg Ile Val Gly His Thr Asp			
130	135	140	
Asn Thr Gly Ser Asp Lys Ile Asn Asp Pro Leu Ser Glu Arg Arg Ala			
145	150	155	160
Ala Ser Val Tyr Ser Phe Leu Asn Ser Gln Gly Val Ser Met Ser Arg			
165	170	175	
Met Ala Ala Glu Gly Arg Gly Ser His Glu Pro Val Ala Asp Asn Ser			
180	185	190	
Thr Val Ala Gly Arg Ser Ala Asn Arg Arg Val Glu Val Tyr Ile L u			
195	200	205	
Pro Asn Ala Lys M t Ile Glu Gln Ala Gln Gln Gly Thr Leu Lys			
210	215	220	

(2) INFORMATION FOR SEQ ID NO:18

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 426 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Porphyromonas gingivalis*

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION 1...426

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18

Met	Glu	Phe	Lys	Ile	Val	Lys	Ile	Asn	Gln	Glu	Tyr	Lys	Asn	Val	Val		
1				5						10					15		
Val	Ser	His	Lys	Val	Leu	Ile	Glu	Ala	Glu	Leu	Glu	Gln	Gln	Lys	Lys		
				20						25					30		
Glu	Ile	Ile	Gly	Lys	Leu	Glu	Lys	Gly	Gln	Val	Leu	Glu	Gly	Ile	Val		
				35				40				45					
Lys	Asn	Ile	Thr	Ser	Tyr	Gly	Val	Phe	Ile	Asp	Leu	Gly	Gly	Val	Asp		
				50				55			60						
Gly	Leu	Ile	His	Ile	Thr	Asp	Leu	Ser	Trp	Gly	Arg	Val	Ala	His	Pro		
				65				70			75				80		
Glu	Glu	Ile	Val	Gln	Leu	Asp	Gln	Lys	Ile	Asn	Val	Val	Ile	Leu	Asp		
				85				90				95					
Phe	Asp	Glu	Asp	Arg	Lys	Arg	Ile	Ala	Leu	Gly	Leu	Lys	Gln	Leu	Met		
				100				105				110					
Pro	His	Pro	Trp	Asp	Ala	Leu	Asp	Ser	Glu	Leu	Lys	Val	Gly	Asp	Lys		
				115				120			125						
Val	Lys	Gly	Lys	Val	Val	Val	Met	Ala	Asp	Tyr	Gly	Ala	Phe	Val	Glu		
				130				135			140						
Ile	Ala	Gln	Gly	Val	Glu	Gly	Leu	Ile	His	Val	Ser	Glu	Met	Ser	Trp		
				145				150			155				160		
Thr	Gln	His	Leu	Arg	Ser	Ala	Gln	Asp	Phe	Leu	His	Val	Gly	Asp	Glu		
				165				170			175						
Val	Glu	Ala	Val	Ile	Leu	Thr	Leu	Asp	Arg	Glu	Glu	Arg	Lys	Met	Ser		
				180				185			190						
Leu	Gly	Leu	Lys	Gln	Leu	Lys	Pro	Asp	Pro	Trp	Ala	Asp	Ile	Glu	Thr		
				195				200			205						
Arg	Phe	Pro	Val	Gly	Ser	Arg	His	His	Ala	Arg	Val	Arg	Asn	Phe	Thr		
				210				215			220						
Asn	Phe	Gly	Val	Phe	Val	Glu	Ile	Glu	Glu	Gly	Val	Asp	Gly	Leu	Ile		
				225				230			235				240		
His	Ile	Ser	Asp	Leu	Ser	Trp	Thr	Lys	Lys	Ile	Lys	His	Pro	Ser	Glu		
				245				250			255						
Phe	Thr	Glu	Val	Gly	Ala	Asp	Ile	Glu	Val	Gln	Val	Ile	Glu	Ile	Asp		
				260				265			270						
Lys	Glu	Asn	Arg	Arg	Leu	S	r	Leu	Gly	His	Lys	Gln	Leu	Glu	Asn		
				275				280			285						
Pro	Trp	Asp	Val	Phe	Glu	Thr	Val	Ph	Thr	Val	Gly	Ser	Ile	His	Glu		
				290				295			300						
Gly	Thr	Val	Il	Glu	Val	M	t	Asp	Lys	Gly	Ala	Val	Val	Ser	L	u	Pro
				305				310			315				320		
Tyr	Gly	Val	Glu	Gly	Ph	Ala	Thr	Pro	Lys	His	Met	Val	Lys	Glu	Asp		
				325				330			335						

Gly Ser Gln Ala Val Leu Glu Glu Lys Leu Pro Phe Lys Val Ile Glu
 340 345 350
 Phe Asn Lys Asp Ala Lys Arg Ile Ile Val Ser His Ser Arg Val Phe
 355 360 365
 Glu Asp Glu Gln Lys Met Ala Gln Arg Glu Ala Asn Ala Glu Arg Lys
 370 375 380
 Ala Glu Ala Lys Ala Ala Gln Lys Glu Ala Ala Ala Glu Ala Ala Asn
 385 390 395 400
 Pro Ala Gln Ala Val Glu Lys Ala Thr Leu Gly Asp Leu Gly Glu Leu
 405 410 415
 Ala Ala Leu Lys Glu Lys Leu Ser Glu Asn
 420 425

References.

1. Mamur, J. (1961) A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *J. Mol. Biol.* 3, 208-218.
2. Fleishmann, R.D. et al. (1995) Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496-512.
3. Borodovsky M, Rudd KE, and EV Koonin. (1994). Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Res.* 22:4756-4767.
4. Horton, P. and Nakai, K. (1996). A probabilistic classification system for predicting the cellular localization sites of proteins. *Intellig. Syst. Mol. Biol.* 4: 109-115.
5. Claros MG and G von Heijne. (1994). TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* 10: 685-686.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Dated this 5th day of May 1998

CSL LIMITED
 Patent Attorneys for the Applicant:
 F.B. RICE & CO.

